

A mathematical model for voice leading and its complexity

Mattia G. Bergomi, Riccardo D. Jadanza, Alessandro Portaluri

25th August 2015

Abstract

We give a formal definition of the musical concept of voice leading in mathematical terms, based on the idea of partial permutations of certain ordered multisets. Then we associate a partial permutation matrix with a voice leading in a unique way and write an algorithm to easily transform any musical composition into a sequence of such matrices; we then generalise it in order to include in the model also rhythmic independence and rests. From that we extract a vector whose components return information about the movements of the voices in the piece and hence about the complexity of the voice leading. We provide some examples by analysing three compositions, also visualising complexity as a point cloud for each piece. Finally, we interpret the sequence of complexity vectors associated with each composition, thus considering the position of each observation with respect to time. The Dynamic Time Warping allows us to compute the distance between two pieces and to show that our approach distinguishes well the examples that we took into account, exhibiting a strong indication that the notion of complexity we propose is a good tool to identify and classify musical pieces.

1 Introduction

In Music, the study of harmony and voice leading concerns the dynamical relation among chords and melodic lines. Chords can be seen as vertical entities supporting a main melodic line, whereas compositional styles can be characterised by the interplay of the various voice lines in horizontal motion. Hence, voice leading can be interpreted as the writing of several melodies interacting together in two ways: voices moving simultaneously affect the listener as chords, whilst the independence in terms of both rhythm and pitch creates a horizontal flow, perceived as a superposition of different themes.

The concurrent motion of two voices is classically referred to as *contrapuntal motion* and is traditionally divided into four classes, illustrated in Figure 3 on page 7. In *contrary motion* the voices move in opposite directions; this gives them contrast and independence [1, Chapter 6], creating an interesting “soundscape” for the listener, as the Canadian composer Raymond M. Schafer calls it. *Parallel*

motion occurs when the interval between the voices is kept constant along the movement; when applied to thirds, sixths and tenths it can be considered among the most powerful voice leading techniques. However, in some cases it impedes the growth in independence of the voices: this is one reason why it is generally avoided when not even forbidden for unisons, octaves and fifths. In *oblique motion* actually only one voice is moving whilst the other remains at the same pitch; this one and *similar motion*, where the two voices proceed in the same direction, convey surely less independence than contrary motion, but undoubtably more than the parallel one.

For practical purposes, voices can be grouped in ranges: from the highest to the lowest, we have:

Soprano :	from C_4 to G_6 ,
Alto :	from G_3 to C_5 ,
Tenor :	from C_3 to G_4 ,
Bass :	from E_2 to C_4 .

Different compositional styles are characterised by different types of motion: the five species of counterpoint represented in Figure 1 arise from various combinations of rhythmic choices and of the amplitudes of the intervals between two consecutive notes. A particular movement of voices is the (*voice*) *crossing*, that occurs when two voices exchange their relative positions – for instance when the Soprano moves below the Alto. This kind of dynamics is generally not desirable because it conveys a sense of discomfort, albeit it is considered less problematic when it involves inner voices (Alto and Tenor) for a few chords.

Given a sequence of chords, an important question is how to transform them into a superposition of voices according to a certain contrapuntal style (see Figure 2). Our aim is to formalise the concept of voice leading in mathematical terms and to build a computationally efficient model for dealing with voice leading based on sparse matrices that encode the motions of the involved voices. This representation allows us to define a notion of *complexity* of the voice leading and to classify different contrapuntal styles, by representing an entire composition as a static point cloud or as a multi-dimensional time series. This second interpretation offers also the possibility to compare two pieces by making use of the *dynamic time warping*, a tool that measures the similarity between them in terms of complexity.

2 Voice leadings, multisets and partial permutations

In general, it is possible to describe a melody as a finite sequence of ordered pairs $(p_i, p_{i+1})_{i \in I}$, where I is a finite set of indices. In order to model the voice leading in a mathematical way it is necessary to introduce first the concept of *multiset*, a generalisation of the idea of set. (This approach was already considered by D. Tymoczko in [5].) Roughly speaking, we can think of it as of a list where an

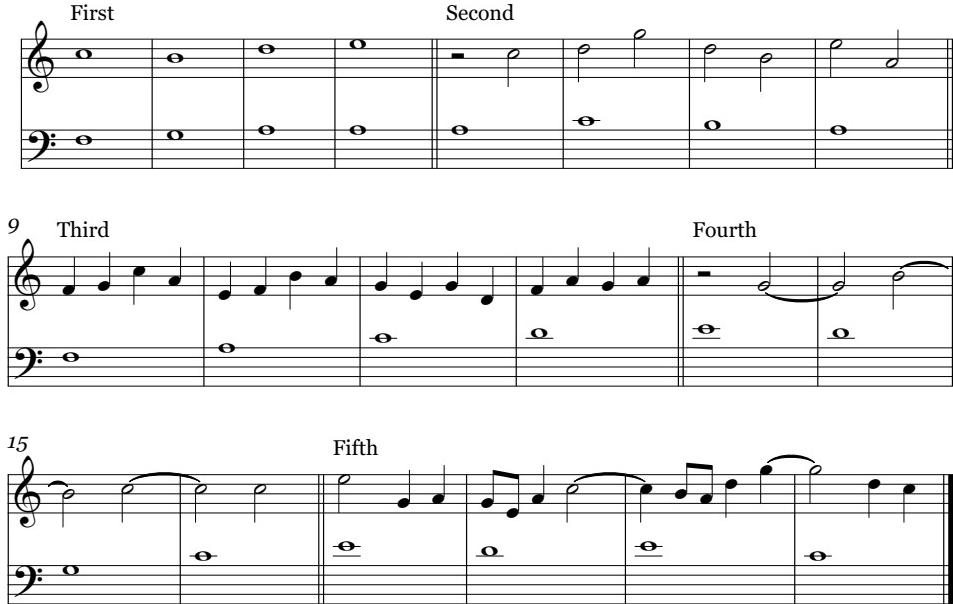


Figure 1: Five different degrees of independence between voices, from the tied one-to-one in the first species of counterpoint to the complete independence of the fifth species.



Figure 2: From a sequence of chords to a superposition of melodies.

object can appear more than once, whilst the elements of a set are necessarily unique. More formally, a *multiset* M is a couple (X, μ) composed of an *underlying set* X and a map $\mu : X \rightarrow \mathbb{N}$, called the *multiplicity* of M , such that for every $x \in X$ the value $\mu(x)$ is the number of times that x appears in M . We define the *cardinality* $|M|$ of M to be the sum of the multiplicities of each element of its underlying set X . Observe, however, that a multiset is in fact completely defined by its multiplicity function: it suffices to set $M := (\text{dom}(\mu), \mu)$.

If we interpret a set of n singing voices (or parts played by n instruments, or both) as a multiset of pitches of cardinality n , then a voice leading can be mathematically described as follows.

Definition 2.1. Let $M := (X_M, \mu_M)$ and $L := (X_L, \mu_L)$ be two multisets of pitches with same cardinality n and arrange their elements into n -tuples (x_1, \dots, x_n) and (y_1, \dots, y_n) respectively.¹ A *voice leading* of n voices between M

¹These are in fact the images of two bijective maps $\psi_M : \{1, \dots, n\} \rightarrow M$ and $\psi_L : \{1, \dots, n\} \rightarrow L$.

and L , denoted by $(x_1, \dots, x_n) \rightarrow (y_1, \dots, y_n)$, is the multiset

$$Z := \{(x_1, y_1), \dots, (x_n, y_n)\},$$

whose underlying set is $X_Z := X_M \times X_L$ and whose multiplicity function μ_Z is defined accordingly, by counting the occurrences of each ordered pair.

Remark 2.2. Observe that the definition just given is not linked to the particular type of object (pitches): it is possible to describe voice leadings also between pitch classes, for instance.

Note that it is also possible to describe a voice leading as a bijective map from the multiset M to the multiset L , i. e. as a partial permutation of the *union multiset*

$$M \cup L := (X_M \cup X_L, \mu_{M \cup L}),$$

where

$$\mu_{M \cup L} := \max\{\mu_M \chi_M, \mu_L \chi_L\}$$

and χ_M and χ_L are the characteristic functions of X_M and X_L , respectively.² We recall that a *partial permutation* of a finite multiset S is a bijection between two subsets of S . In general, if S has cardinality n then this map can be represented as an n -tuple of symbols, some of which are elements of S and some others are indicated by a special symbol — we use \diamond — to be interpreted as a “hole” or an “empty character”. However, since we are not dealing with subsets of a fixed multiset, we shall use the cycle notation to avoid ambiguity and confusion.

Remark 2.3. In order to be able to do computations with partial permutations, it is fundamental to fix an *ordering* among the elements of the union multiset $M \cup L$. We henceforth give $M \cup L$ the natural ordering \leqslant of real numbers, being its elements pitches. Indeed, in classical music with equal temperament, one defines the pitch p of a note as a function of the fundamental frequency ν (measured in Hertz) associated with the sound; more precisely, as the map $p : (0, +\infty) \rightarrow \mathbb{R}$ given by

$$p(\nu) := 69 + 12 \log_2 \left(\frac{\nu}{440} \right).$$

This can be done also in the case where the elements of the union multiset are pitch classes: the ordering is induced by the ordering of their representatives belonging to a same octave. However, in this paper we shall not follow this practice and shall instead restrict to pitches only.

Example 2.4. The voice leading

$$(G_2, G_3, B_3, D_4, F_4) \rightarrow (C_3, G_3, C_4, C_4, E_4) \tag{2.1}$$

²For a multiset S we assume that $\mu_S(x) = 0$ if $x \notin X_S$. With this understanding, the function $\mu_{M \cup L}$ is defined on the whole of $X_M \cup X_L$.

is described by the partial permutation of the ordered union multiset

$$(G_2, C_3, G_3, B_3, C_4, C_4, D_4, E_4, F_4)$$

defined by

$$\begin{pmatrix} G_2 & C_3 & G_3 & B_3 & C_4 & C_4 & D_4 & E_4 & F_4 \\ C_3 & \diamond & G_3 & C_4 & \diamond & \diamond & C_4 & \diamond & E_4 \end{pmatrix}. \quad (2.2)$$

Thus, a voice leading between two multisets of n voices can be seen as a partial permutation of a multiset whose cardinality is less than or equal to $2n$.

The next step is to associate a representation matrix with the partial permutation. Let V be an n -dimensional vector space over a field \mathbb{F} and let $\mathcal{E} := \{e_1, \dots, e_n\}$ be a basis for V . The symmetric group S_n acts on \mathcal{E} by permuting its elements: the corresponding map $S_n \times \mathcal{E} \rightarrow \mathcal{E}$ assigns $(\sigma, e_i) \mapsto e_{\sigma(i)}$ for every $i \in \{1, \dots, n\}$. We consider the well-known *linear representation* $\rho : S_n \rightarrow \text{GL}(n, \mathbb{F})$ of the group S_n given by

$$\rho(1 \ i) := \begin{pmatrix} 0 & & & & 1 & & & \\ & 1 & & & & & & \\ & & \ddots & & & & & \\ & & & 1 & & & & \\ 1 & & & & 0 & & & \\ & & & & & 1 & & \\ & & & & & & \ddots & \\ & & & & & & & 1 \end{pmatrix},$$

where the 1's in the first row and in the first column occupy the positions 1, i and i , 1 respectively. The map ρ sends each 2-cycle of the form $(1 \ i)$ to the corresponding permutation matrix that swaps the first element of the basis \mathcal{E} for the i -th one. Note that each row and each column of a permutation matrix contains exactly one 1 and all its other entries are 0. Following this idea and [2, Definition 3.2.5, p. 165], we say that a matrix $P \in \text{Mat}(m, \mathbb{R})$ is a *partial permutation matrix* if for any row and any column there is at most one non-zero element (equal to 1). When dealing with a voice leading $M \rightarrow L$, the dimension m of the matrix P is equal to the cardinality of the multiset $M \cup L$.

Remark 2.5. In general, the partial permutation matrix associated with a given voice leading is not unique. This is due to the fact that we are dealing with multisets: if $M \rightarrow L$ is a voice leading it is possible that some components of L have the same value, i. e. that different voices are playing or singing the same note.

For this reason we introduce the following convention.

Convention 2.6. *Let $M := (x_1, \dots, x_n) \rightarrow L := (y_1, \dots, y_n)$ be a voice leading and suppose that more than one voice is associated with a same note of L . To this end, let $(x_{i_1}, \dots, x_{i_k})$ be the pitches of M (with $i_1 < \dots < i_k$) that are mapped to the pitches $(y_{j_1}, \dots, y_{j_k})$ of L , with $y_{j_1} = \dots = y_{j_k}$ and $j_1 < \dots < j_k$. In order to uniquely associate a partial permutation matrix $P := (a_{ij})$ with the above voice leading, we assign the value 1 to the corresponding entries of P by following the order of the indices, that is by setting $a_{i_1 j_1} = 1, \dots, a_{i_k j_k} = 1$.*

Thus, we shall henceforth speak of *the* partial permutation matrix associated with a given voice leading.

Example 2.7. The partial permutation matrix associated with the cycle representation (2.2) of voice leading (2.1) is

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Therefore, if $M \rightarrow L$ is a voice leading, if both M and L are thought of as ordered tuples and if P is its partial permutation matrix, we have that $PM = L$; in addition, the “reversed” voice leading $L \rightarrow M$ is obviously described by the transpose P^T of P : $P^T L = M$.

This representation has the advantage of providing objects that are much handier than a multiset of couples, speaking in computational terms. Algorithm 1 presents the pseudocode for the computation of the partial permutation matrix of a voice leading.

Algorithm 1 Computing the partial permutation matrix

Input:

$M \rightarrow L$ ▷ Source (M) and target (L) multisets describing the voice leading

Output:

P ▷ Partial permutation matrix associated with the voice leading

Evaluate multiplicities of all $x \in M$ and all $y \in L$;

Generate the *ordered* multiset $U := M \cup L$;

Initialise $P \in \text{Mat}(|U|, \mathbb{R})$ by setting $P(i, j) = 0$ for all i, j ;

1: **for** $i, j \in \{1, \dots, |U|\}$ **do**

2: **if** $U(i) \rightarrow U(j)$ **then**

3: $P(i, j) = 1$

4: **end if**

5: **end for**



Figure 3: Motion classes for two voices. *Similar*: same direction but different intervals; *parallel*: same direction and same intervals; *oblique*: only one voice is moving; *contrary*: opposite directions.

3 Relative motion of the voices and complexity of a voice leading

We have seen in the previous section how the partial permutation matrix associated with a voice leading contains the information about the passage from one note to the next one for each voice. Here we are going to illustrate that, in fact, the tool that we have built also encodes the direction of motion of the different voices, including the crossings.

On the one hand, in Music one distinguishes between three main behaviours (cf. Figure 3; we omit parallel motion because that is not involved in our analysis):

- *Similar motion*, when the voices move in the same direction;
- *Contrary motion*, when the voices move in opposite directions;
- *Oblique motion*, when only one voice is moving.

On the other hand, with reference to a partial permutation matrix (a_{ij}) , it is possible to describe the motion of a voice by noting three conditions, which are immediate consequences of the ordering of the union multiset:

- 1) If there exists an element $a_{ij} = 1$ for $i < j$ then the i -th voice is moving “upwards”;
- 2) If there exists an element $a_{ij} = 1$ for $i > j$ then the i -th voice is moving “downwards”;
- 3) If there exists an element $a_{ii} = 1$ then the i -th voice is constant.

The connection between the two worlds is the following:

- If either Condition 1) or Condition 2) is verified by two distinct elements then we have *similar motion*;
- If both Condition 1) and Condition 2) hold for two distinct elements then we are facing *contrary motion*;
- The case of *oblique motion* involves Conditions 1) and 3) or Conditions 2) and 3), for at least two distinct elements.

As we mentioned in Section 1, *voice crossing* is a particular case of these motions where the voices swap their relative positions. This phenomenon can be described in terms of multisets as follows.

Definition 3.1. Let $(x_1, \dots, x_n) \rightarrow (y_1, \dots, y_n)$ be a voice leading ($n \in \mathbb{N}$). If there exist two pairs (x_i, y_i) and (x_j, y_j) such that $x_i < x_j$ and $y_i > y_j$ or such that $x_i > x_j$ and $y_i < y_j$ then we say that a *(voice) crossing* occurs between voice i and voice j .

The partial permutation matrix retrieves even this information, as the following proposition shows.

Proposition 3.2. Consider a voice leading of n voices and let $P := (a_{ij})$ be its associated partial permutation matrix. Choose indices $i, j, k, l \in \{1, \dots, n\}$ such that $a_{ij} = 1$ and $a_{kl} = 1$. Then there is a crossing between these two voices if and only if one of the following conditions hold:

- i) $i < k$ and $j > l$;
- ii) $i > k$ and $j < l$.

Furthermore, the total number of voices that cross the one represented by a_{ij} is equal to the number of 1's in the submatrices (a_{rs}) and (a_{tu}) of P determined by the following restrictions on the indices: $r > i$, $s < j$ and $t < i$, $u > j$.

Proof. In a partial permutation matrix the row index of a non-zero entry denotes the initial position of a certain voice in the ordered union multiset, whereas the column index of the same entry represents its final position after the transition. It is then straightforward from Definition 3.1 that for a voice crossing to exist either condition i) or condition ii) must be verified. Every entry a_{kl} satisfying one of those conditions refers to a voice that crosses the one represented by a_{ij} , hence the number of crossings for a_{ij} equals the amount of 1's in positions (r, s) such that $r > i$ and $s < j$, summed to the number of 1's in positions (t, u) such that $t < i$ and $u > j$. \square

Remark 3.3. The fact that the number of crossings with a given voice equals the number of 1's in the submatrices determined by the entry corresponding to that voice (as explained in the previous proposition) holds true only because we assumed Convention 2.6. Indeed, if we did not make such an assumption, the submatrices could contain positive entries referring to voices ending in the same note but that do not produce crossings.

From what we have shown thus far it emerges that it is possible to characterise a voice leading by counting the voices that are moving upwards, those that are moving downwards, those that remain constant and the number of crossings. We summarise these features in a 4-dimensional *complexity vector* c defined by

$$c := (\#\text{upward voices}, \#\text{downward voices}, \#\text{constant voices}, \#\text{crossings}), \quad (3.1)$$

so that we are now able to classify and distinguish voice leadings by simply looking at these four aspects.

Example 3.4. *Similar motion.* The voice leading $(C_1, E_1, G_1) \rightarrow (D_1, F_1, A_1)$ is represented by

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and its complexity vector is $(3, 0, 0, 0)$.

Oblique motion. The voice leading $(G_2, G_2, C_3) \rightarrow (C_3, C_3, C_3)$ is associated with

$$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

and its complexity vector is $(2, 0, 1, 0)$.

Voice crossing. The voice leading $(C_1, E_1, G_1) \rightarrow (G_1, C_1, E_1)$ is represented by

$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

and its complexity vector is $(1, 2, 0, 2)$.

By virtue of these tools it is straightforward to analyse an entire piece of music: it is enough to divide it into pairs of notes for each voice and apply the procedure described above for each passage. The concatenation of all the consecutive passages results then in a sequence of partial permutation matrices, whence one can extract a sequence of complexity vectors. This last piece of information can be visualised as a set of points in a 4-dimensional space — or rather as one or more of its 3-dimensional projections (see Subsection 3.1). In fact, if one wants to represent the complexity of the whole composition as a point cloud, one should take into account that different matrices can produce the same complexity vector; therefore we have a *multiset* of points in \mathbb{R}^4 (with non-negative integer components).

3.1 Complexity analysis of two *Chartres Fragments*

We are going to analyse two pieces that are parts of the *Chartres Fragments*, an ensemble of compositions dating back to the Middle Ages: *Angelus Domini* and *Dicant nunc Judei*; both of them are counterpoints of the first species and

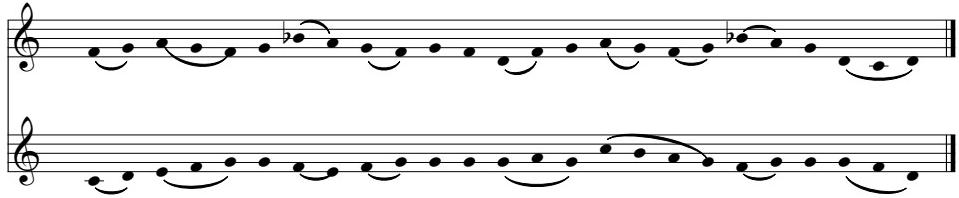


Figure 4: *Angelus Domini*

involve only two voices. The musical interest in these compositions consists in the introduction of a certain degree of independence between the voices and the use of a *parsimonious voice leading*, i. e. an attempt to make the passage from a melodic state to the next as smooth as possible. Note how the independence of the voices is reflected by the presence of contrary motions and crossings, which can then be interpreted as a rough measure of this feature. For a complete treatise on polyphony and a historical overview we refer the reader to [4].

In what follows, we represent the multiplicity of each complexity vector c as a circle of centre $c \in \mathbb{R}^4$ and radius equal to the *normalised multiplicity* $\mu(c)/n$ of c , where $\mu(c)$ is the number of occurrences of c in the analysed piece and n is the total number of notes played or sung by each voice in the whole piece.

Angelus Domini. The fragment under examination is depicted in Figure 4; here is the list of its first four voice leadings, as they are generated by the pseudocode described in Algorithm 1:

```
Voice Leading: ['F4', 'C4'] ['G4', 'D4']
[2, 0, 0, 0] - similar motion up
Voice Leading: ['G4', 'D4'] ['A4', 'E4']
[2, 0, 0, 0] - similar motion up
Voice Leading: ['A4', 'E4'] ['G4', 'F4']
[1, 1, 0, 0] - contrary motion
Voice Leading: ['G4', 'F4'] ['F4', 'G4']
[1, 1, 0, 1] - contrary motion - 1 crossing
```

Table 1 on page 12 contains the the complexity vectors and their occurrences in the piece; the point cloud associated with this multiset is represented in Figure 5. Observe how the projection that neglects the component of c corresponding to the number of constant voices (Figure 5b) gives an immediate insight on the relevance of voice crossing in the piece.

Dicant nunc Judei. The first part of the output of Algorithm 1 produces the following analysis:

```
Voice Leading: ['F4', 'C4'] ['G4', 'E4']
```

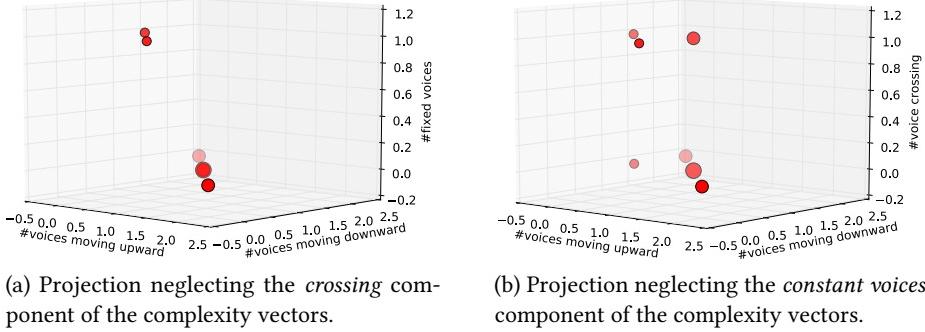


Figure 5: Three-dimensional projections of the complexity cloud of the paradigmatic voice leading *Angelus Domini*. The radius of each circle represents the normalised multiplicity of the corresponding complexity vector.

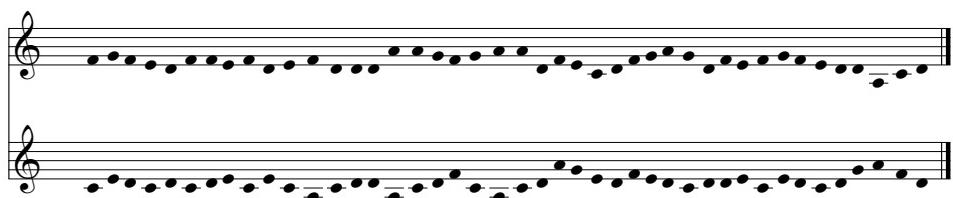


Figure 6: *Dicant nunc Judei*, Chartres fragment.

```
[2, 0, 0, 0] - similar motion up
Voice Leading: ['G4', 'E4'] ['F4', 'D4']
[0, 2, 0, 0] - similar motion down
Voice Leading: ['F4', 'D4'] ['E4', 'C4']
[0, 2, 0, 0] - similar motion down
Voice Leading: ['E4', 'C4'] ['D4', 'D4']
[1, 1, 0, 1] - contrary motion - 1 crossing
```

The complexity vectors arising in the whole piece and their multiplicities are again collected in Table 1; see Figure 7 instead for a visualisation of the point cloud describing the piece. Note how the voice crossing is more massive than in the point cloud describing *Angelus Domini*. In addition, the point $(0, 0, 0)$ in Figure 7b corresponds to the point $(0, 0, 2, 0) \in \mathbb{R}^4$, that represents trivial voice leadings where both parts do not vary.

4 Rhythmic independence and rests

The examples analysed in Subsection 3.1 are counterpoints of the first species – which is the simplest case, in that the voices follow a *note-against-note* flow. It is however possible to study more complex scenarios by introducing *rhythmic*

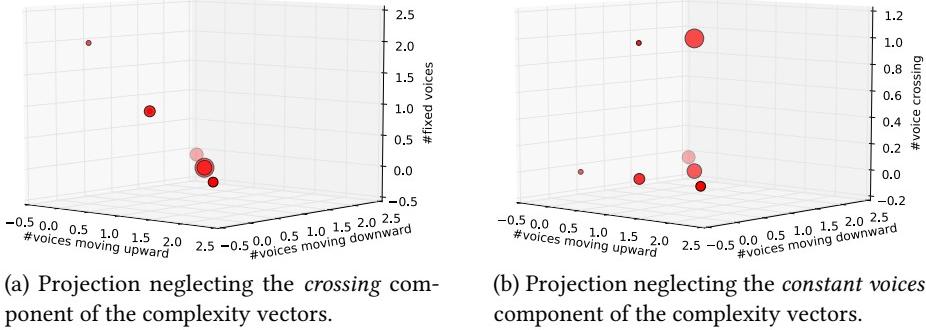


Figure 7: Three-dimensional projections of the complexity cloud of the paradigmatic voice leading *Dicant nunc Judei*. The radius of each circle represents the normalised multiplicity of the corresponding complexity vector.

Table 1: Complexity vectors of the analysed fragments and their occurrences.

<i>Angelus Domini</i>		<i>Dicant nunc Judei</i>	
c	$\mu(c)$	c	$\mu(c)$
(0, 1, 1, 0)	2	(0, 0, 2, 0)	1
(0, 1, 1, 1)	2	(0, 2, 0, 0)	7
(0, 2, 0, 0)	4	(1, 0, 1, 0)	5
(1, 0, 1, 1)	2	(1, 0, 1, 1)	1
(1, 1, 0, 0)	6	(1, 1, 0, 0)	9
(1, 1, 0, 1)	4	(1, 1, 0, 1)	15
(2, 0, 0, 0)	4	(2, 0, 0, 0)	4

independence between voices and *rests* in the melody, in any case reducing non-simultaneous voices to the simplest case.

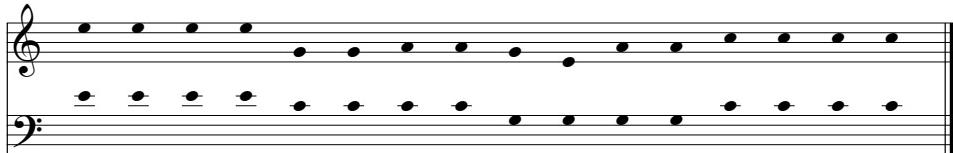
If the voices play at different rhythms or follow rhythmically irregular themes, we consider the minimal rhythmic unit u appearing in the phrase and *homogenise* the composition based on that unit: if a note has duration ku , with $k \in \mathbb{N}$, we represent it as k repeated notes of duration u (see Figure 8 for an example). This transformation of the original counterpoint introduces only oblique motions and does not alter the number of the other three kinds of motion.

In musical terms, if a voice is silent it is neither moving nor being constant and it cannot cross other voices. Therefore, in order to include rests in our model it is necessary to slightly modify Algorithm 1 by introducing a new symbol (p) in the dictionary of pitches; we also choose to indicate a rest in the matrices associated with a voice leading by the entry -1 . We adopt the following convention concerning the ordered union multiset.

Convention 4.1. *We choose rests to be the last elements in the ordered union multiset associated with a voice leading. In other words, we declare p to be strictly*



(a) Counterpoint of the fifth species.



(b) Reduction to the first species.

Figure 8: Reduction of rhythmically independent voices to a counterpoint of the first species.

greater than any other pitch symbol.

Example 4.2. The voice leading $(p, D_4, D_5) \rightarrow (D_4, C_3, C_3)$ corresponds to the matrix

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{pmatrix}.$$

Remark 4.3. Note that when introducing the -1 's in the matrix associated with a voice leading we are no longer dealing with partial permutation matrices. However, to study voice leadings with rhythmic independence of the voices as before (thus ignoring rests) it is enough to consider the minor of the matrix obtained by deleting all rows and columns containing -1 (which is obviously again a partial permutation matrix).

We extend the complexity vector defined previously in Formula (3.1) by adding a fifth component that counts the number of voices that are silent at least once in the voice leading, i. e. it counts the number of negative (-1) entries of the associated matrix. Furthermore, we slightly modify also the notion of *normalised multiplicity* of a complexity vector c , needed for the representation of the complexity of a piece in the form of a point cloud, now dividing the number $\mu(c)$ of occurrences of c in the piece by the total number of notes per voice *after the homogenisation*.

4.1 Example: the *Retrograde Canon* by J. S. Bach

We consider the *Retrograde Canon* (also known as *Crab Canon*), a palindromic canon with two voices belonging to the *Musikalisches Opfer* by J. S. Bach, the



Figure 9: The *Retrograde Canon* (bars 1–4), a palindromic canon belonging to the *Musikalisches Opfer* by J. S. Bach.

Table 2: Complexity vectors of the *Retrograde Canon* and their occurrences.

Retrograde Canon			
c	$\mu(c)$	c	$\mu(c)$
(0, 0, 1, 0, 1)	2	(1, 0, 0, 0, 1)	2
(0, 0, 2, 0, 0)	8	(1, 0, 1, 0, 0)	43
(0, 1, 0, 0, 1)	2	(1, 0, 1, 1, 0)	1
(0, 1, 1, 0, 0)	43	(1, 1, 0, 0, 0)	14
(0, 1, 1, 1, 0)	1	(1, 1, 0, 1, 0)	3
(0, 2, 0, 0, 0)	11	(2, 0, 0, 0, 0)	11

beginning of which is reproduced in Figure 9.

We homogenise the rhythm by expressing each note in eighths and we apply Algorithm 1. Here is the output of the first four meaningful voice leadings:

```

Voice Leading: ['D4', 'D4'] ['D4', 'F4']
c = [1, 0, 1, 0, 0] - oblique motion
Voice Leading: ['D4', 'F4'] ['F4', 'A4']
c = [2, 0, 0, 0, 0] - similar motion up
Voice Leading: ['F4', 'A4'] ['F4', 'D5']
c = [1, 0, 1, 0, 0] - oblique motion
Voice Leading: ['F4', 'D5'] ['A4', 'C#5']
c = [1, 1, 0, 0, 0] - contrary motion

```

Table 2 collects the complexity vectors and their multiplicities; they are displayed in the form of point clouds in Figure 10.

5 Concatenation of voice leadings and time series

The paradigmatic point cloud associated with a voice leading gives a useful 3-dimensional representation of the piece; however, this analysis is just structural, as it does not take into account the way in which voice leadings have been concatenated by the composer. It is possible to introduce this temporal dimension by looking at the sequence of complexity vectors from a different viewpoint.

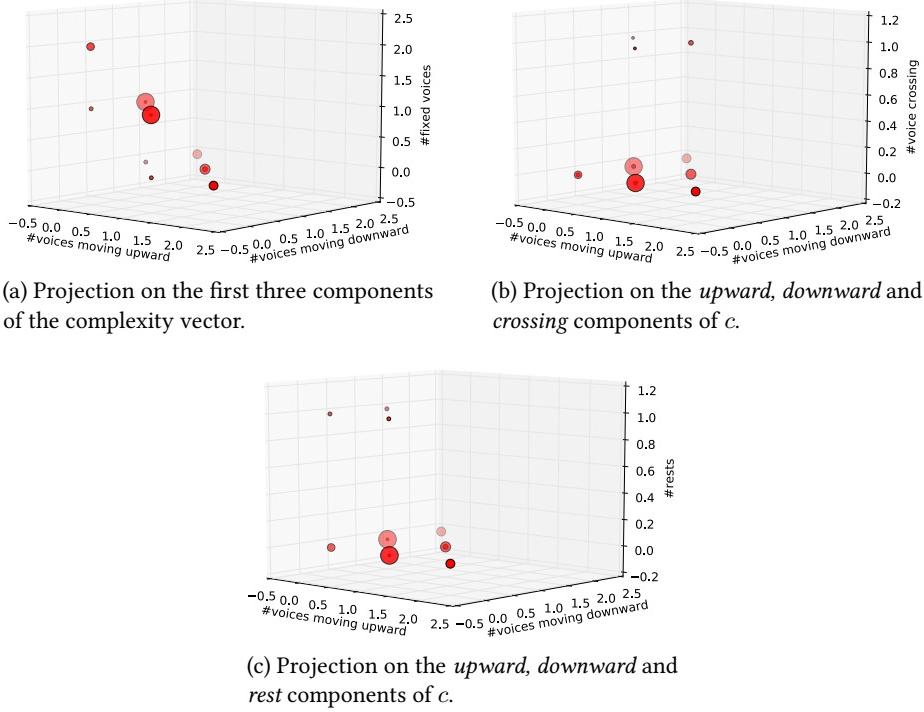


Figure 10: Three-dimensional projection of the 5-dimensional point cloud representing the complexity of the *Retrograde Canon*. The radius of each circle represents the normalised multiplicity of each complexity vector.

The concatenation of observations in time can be seen as a *time series*, that is a sequence of data concerning observations ordered according to time. In our case each piece of music can be described as a 5-dimensional time series, whose observations are the complexity vectors associated with each voice leading. More specifically, we use the so-called *dynamic time warping (DTW)*, a method for comparing time-dependent sequences of different lengths: it returns a measure of similarity between two given sequences by “warping” them non-linearly (see Figure 11 for an intuitive representation). We invite the reader to consult [3] for a detailed review of DTW algorithms.

5.1 Dynamic time warping analysis

Let \mathcal{F} be a set, called the *feature space*, and take two finite sequences $X := (x_1, \dots, x_n)$ and $Y := (y_1, \dots, y_m)$ of elements of \mathcal{F} , called *features* (here n and m are natural numbers). In order to compare them, we need to introduce a notion of distance between features, that is a map $\mathcal{C} : \mathcal{F} \times \mathcal{F} \rightarrow \mathbb{R}$, also called a *cost function*, that meets at least the following requirements:

- i. $\mathcal{C}(x, y) \geq 0$ for all $x, y \in \mathcal{F}$;

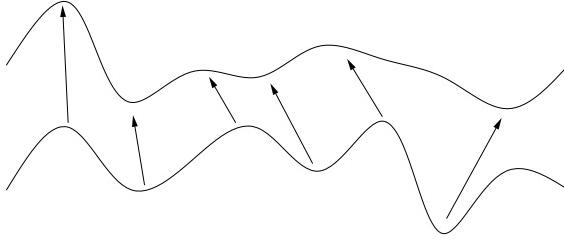


Figure 11: Comparing two time series with DTW.

- ii. $\mathcal{C}(x, y) = 0$ if and only if $x = y$;
- iii. $\mathcal{C}(x, y) = \mathcal{C}(y, x)$ for all $x, y \in \mathcal{F}$.

Now, if we apply \mathcal{C} to the features X and Y , we can arrange the values in an $n \times m$ real matrix $C := (\mathcal{C}(x_i, y_j))$, where i ranges in $\{1, \dots, n\}$ and j in $\{1, \dots, m\}$.

A (n, m) -warping path in C is a finite sequence $\gamma := (\gamma_1, \dots, \gamma_l) \in \mathbb{R}^l$, with $l \in \mathbb{N}$, such that:

1. $\gamma_k := (\gamma_k^x, \gamma_k^y) \in \{1, \dots, n\} \times \{1, \dots, m\}$ for all $k \in \{1, \dots, l\}$;
2. $\gamma_1 := (1, 1)$ and $\gamma_l := (n, m)$;
3. $\gamma_k^x \leq \gamma_{k+1}^x$ and $\gamma_k^y \leq \gamma_{k+1}^y$ for all $k \in \{1, \dots, l - 1\}$;
4. $\gamma_{k+1} - \gamma_k \in \{(1, 0), (0, 1), (1, 1)\}$ for all $k \in \{1, \dots, l - 1\}$.

The *total cost* of a (n, m) -warping path γ over the features X and Y is defined as

$$\mathcal{C}_\gamma(X, Y) := \sum_{k=1}^l \mathcal{C}(x_{\gamma_k^x}, y_{\gamma_k^y}).$$

An *optimal warping path* on X and Y is a warping path realising the minimum total cost (see Figure 12). We are now ready to define the *DTW distance* between X and Y :

$$DTW(X, Y) := \min \{ \mathcal{C}_\gamma(X, Y) \mid \gamma \text{ is a } (n, m)\text{-warping path} \}.$$

Remark 5.1. Note that the minimum always exists because the set is finite.

We computed the DTW distance between each pair of the three examples that we analysed in Subsections 3.1 and 4.1, choosing as cost function the Euclidean distance in \mathbb{R}^5 . We embedded the 4-dimensional complexity vectors in \mathbb{R}^5 by adding a fifth component and setting it to 0. The results of the comparison are shown in Table 3. Although we analysed only three compositions, it is possible to observe how the DTW distance segregates the two pieces belonging to the Chartres fragments.

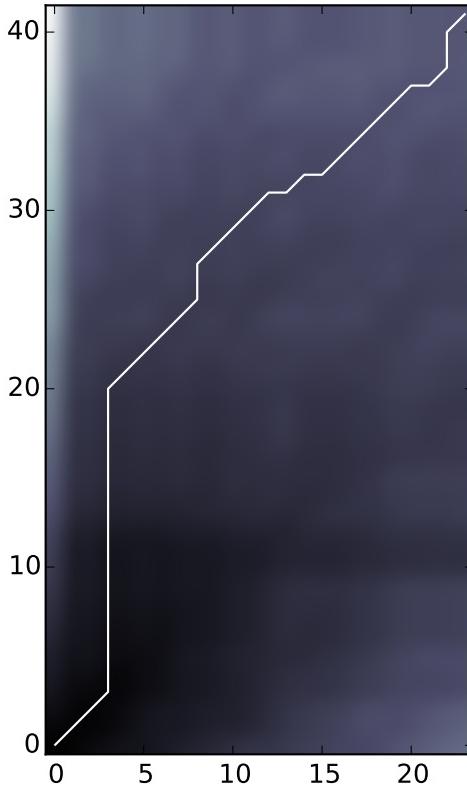


Figure 12: Optimal warping path on *Angelus Domini* and *Dicant nunc Judei*.

6 Conclusion

Our analysis showed that our definition of complexity in terms of the relative movements of the voices and especially of crossing is suitable for characterising a musical piece. Point-cloud representation yields a “photograph” of complexity, a sort of fingerprint that lets clearly emerge what are the main features of the examined composition, noticeable even at first glance. Dynamic time warping provides then further support to this evidence by directly measuring the distance between the complexities of two pieces, considering each complexity vector as an observation of the piece in time, and giving a quantitative description of the

Table 3: DTW distance matrix for the three time series of complexity vectors.

	<i>Angelus</i>	<i>Dicant</i>	<i>Canon</i>
<i>Angelus</i>	0.00	0.62	1.34
<i>Dicant</i>	0.62	0.00	1.16
<i>Canon</i>	1.34	1.16	0.00

dissimilarity of the time series describing the pieces.

References

- [1] E. Aldwell, C. Schachter, and A. Cadwallader. *Harmony and voice leading*. Cengage Learning, 2010.
- [2] R. A. Horn and C. R. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, 1991.
- [3] P. Senin. Dynamic time warping algorithm review. *University of Hawaii*, 2008.
- [4] R. Taruskin. *Music in the Nineteenth Century: The Oxford History of Western Music*. Oxford University Press, 2009.
- [5] D. Tymoczko. The geometry of musical chords. *Science*, (313):72–74, 2006.